REVISTA SIGNOS. ESTUDIOS DE LINGÜÍSTICA ISSN 0718-0934
© 2026 PUCV, Chile 59(122) 1-0

El impacto de la inteligencia artificial en la traducción jurídica: un análisis multidimensional[1]

The Impact of Generative Models on Legal Translation: A Multidimensional Analysis

María del Mar Sánchez Ramos
UNIVERSIDAD DE ALCALÁ
ESPAÑA
mar.sanchezr@uah.es

Muhammad Shakir
UNIVERSITY OF MÜNSTER
ALEMANIA
muhammad.shakir@uni-muenster.de

Recibido: 24-11-2025 / **Aceptado**: 9-9-2026
DOI: 10.4151/S0718-09342026012201394

Resumen

Este estudio exploratorio aplica el análisis multidimensional de Biber (1988) con el fin de examinar las características lingüísticas de la traducción humana (TH), la traducción automática neuronal (TAN) y la traducción generada por un modelo de IA (GPT-4) en el ámbito jurídico, concretamente en sentencias del Tribunal de Justicia de la Unión Europea (TJUE). Aunque frecuente en lingüística de corpus, su aplicación a los estudios de traducción sigue siendo incipiente. A partir de un corpus de más de 2,6 millones de palabras etiquetado con MFTE, el estudio aplica el análisis multidimensional para identificar patrones lingüísticos coocurrentes en los tres métodos de traducción. Los resultados revelan diferencias significativas, especialmente en dos dimensiones: una relacionada con la complejidad expositiva, y otra con el lenguaje procedimental. Las traducciones de GPT-4 muestran una mayor densidad sintáctica y expresión procedimental, en consonancia con las convenciones estilísticas del discurso jurídico. Por el contrario, los resultados de TAN tienden a ser más abstractos conceptualmente y menos variados estructuralmente, mientras que TH muestra un equilibrio entre la formalidad y la adaptación comunicativa. Estos hallazgos subrayan el valor de este enfoque metodológico y ponen de relieve la variación estilística y funcional entre los métodos de traducción examinados. El estudio ofrece implicaciones metodológicas y pedagógicas, y abre camino a futuras investigaciones sobre la calidad y la variación de la traducción automática.

Palabras clave: variación lingüística, modelos generativos de traducción, sentencias del TJUE, corpus paralelo, traducción automática neuronal

Abstract

This exploratory study applies Biber's (1988) multidimensional analysis (MDA) to examine the linguistic features of human translation (HT), neural machine translation (NMT), and translation produced by a generative AI model (GPT-4) within the legal domain, specifically in judgments issued by the Court of Justice of the European Union (CJEU). While MDA has been widely adopted in corpus linguistics, its application to translation studies remains incipient. Drawing on a corpus of over 2.6 million words, the study applies multidimensional analysis, using MFTE-tagged texts, to identify co-occurring linguistic patterns across the three translation methods. The results reveal significant differences, particularly in two key dimensions: one related to syntactic complexity and formal exposition, and another to procedural and event-focused language. GPT-4 translations exhibit higher syntactic density and procedural expression, consistent with the stylistic conventions of legal discourse. In contrast, NMT outputs tend to be more conceptually abstract and less structurally varied, while HT demonstrates a balance between formality and communicative adaptation. These findings underscore the value of this methodological approach and highlight the stylistic and functional variation associated with the different translation methods examined. The study offers both methodological and pedagogical implications, paving the way for future research into machine translation quality and variation.

Keywords: linguistic variation, generative translation models, CJEU judgments, parallel corpus, neural machine translation

INTRODUCCIÓN

Avanzar en la caracterización de las distintas manifestaciones del lenguaje es imprescindible no solo para generar conocimiento lingüístico teórico y descriptivo, sino también para las aplicaciones de la lingüística computacional, el procesamiento del lenguaje natural o la traducción automática (TA). La llegada de la inteligencia artificial (IA) generativa ha supuesto un avance sin precedentes en la traducción: esta tecnología, materializada en modelos de lenguaje masivo permite traducir y revisar textos mediante los llamados *prompts*. Antes de la IA generativa, la investigación sobre traducción automática neuronal (TAN) se centraba en el llamado *machine translationese* y el *post-editese*. No obstante, estos trabajos priorizan la calidad de los textos finales, dejando un vacío investigador sobre el comportamiento lingüístico de la TA y la IA generativa frente a la TH (Toral, 2019; Brivá-Iglesias et al., 2024).

Con el objetivo de abrir vías de investigación en TA, nuestro trabajo contribuye a los estudios sobre TA desde una perspectiva cuantitativa y exploratoria y proponemos el análisis multidimensional (Biber, 1988) para caracterizar la TH, la TAN y GPT-4, identificar sus rasgos lingüísticos coocurrentes y establecer las dimensiones funcionales que distinguen estos tres métodos de traducción. Este enfoque cuantitativo, apoyado en el análisis estadístico de un corpus amplio, permite identificar patrones lingüísticos funcionales y regularidades discursivas que complementan los hallazgos de las aproximaciones cualitativas, tradicionalmente predominantes en la investigación sobre traducción jurídica. Nuestra hipótesis inicial es que la interacción entre la TH, la TAN

y GPT-4 queda reflejada en las características léxico-gramaticales de los textos traducidos.

1. El avance de la IA: de lo humano a lo automático

Las últimas décadas han destacado por el interés que la traducción ha despertado como proceso de comunicación. Los estudios de traducción basados en corpus han proporcionado nuevas vías de investigación, como son los trabajos centrados en la búsqueda de patrones léxicos "universales" —el uso de un lenguaje simplificado y conservador, el alto nivel de explicitación o el alto grado de homogenización— han puesto de manifiesto las características de la TH (Baker, 1993; Ilisei & Inkpen, 2011; Volansky et al., 2015). Bajo este nuevo paradigma se incluyen trabajos que comparan la TH y la TA en busca de patrones universales traductológicos, como los de Lapshinova-Koltunski (2015) —simplificación, explicitación y convergencia entre TH y dos tipos de TA—, Krüger (2020) —explicitación léxica en la TAN— o Rico Pérez (2025) —efecto de imprimación de la TA en textos institucionales de la UE traducidos al español—. En torno al llamado *post-editese*, Toral (2019) documenta baja densidad léxica en textos poseditados de distintas lenguas, en la línea de la pérdida de riqueza léxica que cuantifican, de forma más general Vanmassenhove et al. (2019), mientras que Castilho y Resende (2022) exploran el fenómeno en el ámbito literario. Desde una perspectiva más lingüística, Frankenberg-Garcia (2022) confirma, en el par portugués-inglés y en textos literarios, una inconsistencia léxica considerable en la TAN, y Niu y Jiang (2024), con un corpus comparable de distintos géneros, concluyen que la simplificación es la principal característica de la huella de la TA.

Si nos detenemos en la traducción jurídica, los textos producidos en el marco institucional de la Unión Europea (UE) cuentan con una sólida tradición de estudios de corpus, entre los que ocupa un lugar central el análisis del Eurolecto —la variedad híbrida que emerge del proceso de traducción en contextos institucionales multilingües— (Biel, 2014). La investigación sobre TA en este ámbito cuenta con trabajos consolidados centrados en evaluar su calidad (Heiss & Soffritti, 2018; Killman, 2023; Yates, 2006) con resultados que van del rechazo inicial de Yates (2006) a su consideración como "prometedora" en Heiss y Soffritti (2018), lo que indica una implantación progresiva en distintos campos. Esta evolución se observa también en la práctica profesional. Tal es el caso de Góngora-Goloubintseff (2026), que muestra, a partir de una encuesta a traductores jurídicos autónomos del Reino Unido, que la adopción de la TAN es desigual —más extendida entre los profesionales de incorporación más reciente— y que la posedición humana sigue considerándose imprescindible. Por último, Brivá-Iglesias et al. (2024) evalúan, mediante métricas automáticas, la TAN (*Google Translate*) y los modelos de lenguaje masivo (GPT-4 y Vicuna) en varias lenguas, y Giampieri (2025) hace lo propio para el par inglés-italiano. Estos trabajos confirman que la aplicación de la IA en contextos jurídicos es una línea

de investigación activa cuyo estudio puede ayudar a conocer y mejorar su aplicación potencial.

2. El análisis multidimensional: aplicaciones recientes en los estudios de traducción

Propuesto por Biber (1988) y extendido en sucesivas publicaciones (Biber, 1995a, 1995b), el análisis multidimensional surgió para describir la variación lingüística entre los registros de inglés hablado y escrito. El enfoque parte de que la variación entre registros no se explica por un solo parámetro, sino que existe un continuum de variación lingüística a lo largo de los registros. De ahí uno de sus principios fundamentales: la coocurrencia de rasgos lingüísticos (léxicos, gramaticales, sintácticos y retóricos), determinada mediante procedimientos estadísticos avanzados (p. ej., el análisis factorial), que explicaría su presencia en los textos debido a funciones comunes (situacionales, sociales y cognitivas). Esta coocurrencia es lo que se conoce como dimensiones. Las dimensiones establecidas por Biber (1988) siguen vigentes en la actualidad como indicadores de variación en inglés (Berber-Sardinha, 2024.)

Conviene distinguir dos aproximaciones metodológicas al aplicar el análisis multidimensional. Los llamados *additive multidimensional analyses* se identifican con los análisis factoriales confirmatorios y comparan el estudio con dimensiones ya establecidas, normalmente las de Biber (1988), sin requerir un nuevo análisis factorial. El llamado "novel multidimensional" o *full MDA*, en cambio, no parte de dimensiones previas: es inductivo y extrae e interpreta factores o componentes (dimensiones) específicos mediante técnicas de reducción de dimensionalidad como el análisis factorial exploratorio o el análisis de componentes principales (ACP). Ambas técnicas identifican las correlaciones entre las variables y agrupan las que se distribuyen de forma similar en un mismo factor o componente, interpretado después en términos de categorías funcionales como una dimensión de variación.

Los estudios de traducción no han sido ajenos a la influencia del análisis multidimensional, aunque a un ritmo lento. Tal es el caso de De Sutter y Lefer (2020), quienes abordan la traducción como una actividad y un producto lingüístico intrínsecamente multidimensional y proponen lo que llaman "a new, updated research agenda" para los estudios de traducción basados en corpus. Destaca también la propuesta de un nuevo análisis multidimensional de Hu et al. (2016) para comparar el inglés escrito traducido y no traducido. En el discurso jurídico-institucional de la UE, Biel et al. (2025) aplican también un nuevo análisis multidimensional a cuatro registros institucionales en polaco —incluidas las sentencias del TJUE—, mientras que Sánchez Ramos y Shakir (2026) llevan a cabo este mismo análisis en un corpus de traducciones humanas y automáticas de sentencias del TJUE al inglés, identificando dos dimensiones con elevado poder diferenciador entre ambos métodos. El presente artículo amplía ese diseño al incorporar la producción de IA generativa (GPT-4) como tercera condición

de análisis. Entre los estudios que comparan sus resultados con las dimensiones de Biber —los análisis factoriales confirmatorios— destacan Kruger & Van Rooy (2016), quienes formulan la hipótesis del lenguaje restringido: la traducción y las variedades autóctonas no nativas del inglés, producidas bajo condiciones de contacto lingüístico, comparten rasgos de explicitación, simplificación y homogeneización que se traducen en perfiles menos marcados en las dimensiones de Biber. Otros trabajos son los de Chou y Liu (2024), centrado en la variación intratextual de la traducción de ficción; o Calzada Pérez y Sánchez Ramos (2022), quienes comprueban, también de forma confirmatoria, si dichas dimensiones son aplicables a textos traducidos y no traducidos de distintas instituciones parlamentarias.

3. Metodología

La investigación que aquí se presenta aborda un nuevo análisis multidimensional para acometer un estudio de naturaleza cuantitativa y medir el grado de similitud y diferencia entre la TH, la TAN y GPT-4 en el ámbito jurídico, más concretamente en el género de las sentencias emitidas por el TJUE. El corpus empleado es un subconjunto del corpus paralelo inglés-español del TJUE (Vigier Moreno & Sánchez Ramos, 2017). Las TH son sentencias emitidas por el TJUE durante el periodo 1990-2014 traducidas del español al inglés. En cuanto a la TAN, los textos fuente en español fueron traducidos por MTUOC-Translator (Oliver, 2021), un motor neuronal entrenado con el Corpus Paralelo de las Naciones Unidas (Ziemski et al., 2016).[2] Este conjunto de datos de entrenamiento es uno de los mayores existentes de libre acceso y se ha utilizado en otros estudios sobre evaluación de TA (Rico Pérez et al., 2020). GPT-4, por su parte, se entrena con datos de acceso propietario cuya composición exacta no se conoce públicamente y cuyo volumen y diversidad son cualitativamente distintos a los de MTUOC. En cuanto a GPT-4, los textos se obtuvieron a través de la API del modelo. Como modelo de lenguaje masivo, GPT-4 necesita de los llamados *prompts* para generar una traducción. El objetivo de nuestro estudio no era optimizar la calidad de la traducción mediante técnicas de ingeniería de *prompts*, sino analizar el comportamiento lingüístico del sistema en condiciones controladas y comparables con las otras dos modalidades de traducción. Con este fin, se optó deliberadamente por un *prompt* mínimo de tipo *zero-shot* orientado exclusivamente a la tarea (*Translate the following text into English*), aplicado de manera uniforme a todos los textos del corpus. Este tipo de *prompt* ha sido empleado en investigaciones comparables en traducción especializada (Brivá-Iglesias et al., 2024; Jiao et al., 2023). La incorporación de instrucciones adicionales —relativas al rol del traductor, al dominio jurídico, al estilo o a la terminología— habría introducido variables externas capaces de modificar precisamente los rasgos léxico-gramaticales que el análisis multidimensional pretende capturar, introduciendo un factor de variación adicional que dificultaría la comparabilidad interna del corpus. Dicho *prompt* se generó con fecha de abril 2025 en GPT-4. Con ello, el conjunto de datos empleado, y que constituyó el corpus final de trabajo, incluye TH, TAN y GPT-4 (Tabla 1).[3]

Tabla 1

Datos del corpus final

Tipo de traducción	Número de palabras	Número de archivos
TH	864.464	158
TAN	866.456	158
GPT-4	889.796	158
Total	2.620.716	474

Los archivos de texto se etiquetaron con MFTE, una herramienta que identifica y etiqueta varios rasgos léxico-gramaticales relevantes para estudios que apliquen el análisis multidimensional (Le Foll & Shakir, 2023, 2024). Los rasgos etiquetados por MFTE incluyen los utilizados en el estudio original de Biber (1988) y las etiquetas ampliadas añadidas a partir de sus trabajos posteriores (Biber, 2006; Biber et al., 1999). La versión 1.6 utilizada en este análisis incorpora además seis rasgos adicionales a nivel de oración y de sintagma, lo que aproxima su resultado al conjunto de rasgos habitualmente utilizado en la mayoría de los estudios de análisis multidimensional. Atendiendo a lo establecido por Biber (1988), nuestro análisis emplea frecuencias normalizadas por cada 100 palabras. Basándonos en recomendaciones anteriores, se eliminaron los rasgos con correlaciones muy altas ($>|0{,}9|$) o muy bajas ($<|0{,}1|$) (Le Foll, 2022) y los rasgos con bajas coincidencias ($< 0{,}2$) siguiendo a Egbert y Staples (2019). Además, se eliminó cualquier característica con un valor Kaiser-Meyer-Olkin (KMO) inferior a 0,5 para mejorar la factorabilidad de los datos (Le Foll, 2022). Para reducir el impacto de los valores atípicos, los datos se transformaron utilizando una puntuación z seguida de una transformación logarítmica con signo, como sugieren Neumann y Evert (2021). Según estos autores, esta transformación contribuye a reducir el impacto de los valores atípicos. Según nuestras propias observaciones, reduce además el riesgo de que un mismo texto figure entre los de puntuación más alta (es decir, entre los diez primeros) en varias dimensiones a la vez.

Uno de los primeros pasos en el análisis multidimensional consiste en determinar el número de factores o componentes que se van a extraer, para lo cual existen varios métodos. El gráfico *scree plot* (Figura 1) es uno de ellos. Como ilustra la Figura 1, los valores propios de los datos observados tienden a estabilizarse a partir del sexto componente. Los resultados obtenidos mediante análisis paralelo a partir de datos simulados resultan algo más claros para el análisis de componentes principales (ACP) que para el análisis factorial exploratorio (AFE). En ambos casos, no obstante, el análisis paralelo sugiere la extracción de diez o más factores o componentes principales (CP). El test MAP, aplicado mediante la función psych:VSS para determinar el número óptimo de factores, sugiere un mínimo de quince. Si bien la extracción de diez o más factores o CP puede resultar estadísticamente razonable, desde el punto de vista lingüístico resulta cada vez más difícil interpretar componentes integrados por solo dos o unos pocos rasgos lingüísticos. Por ello, se han seleccionado los seis primeros

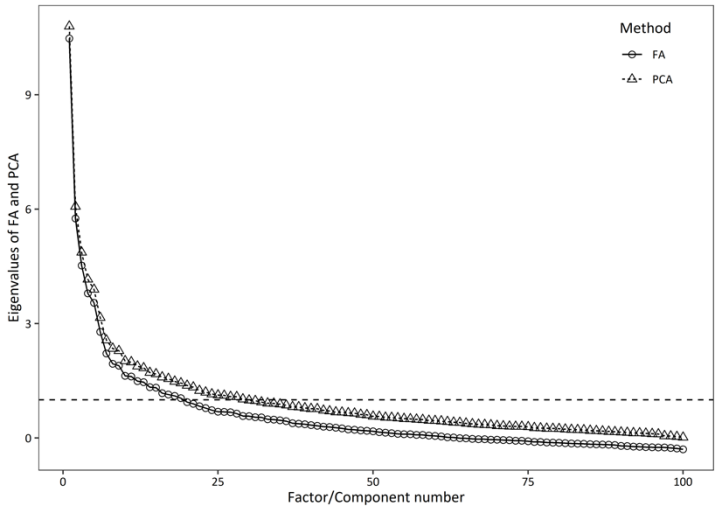
componentes, tal como indica el *scree plot* y atendiendo a su interpretabilidad en estos datos.

El ACP es una técnica de reducción de dimensionalidad afín al AFE y, como este, se emplea habitualmente en estudios de análisis multidimensional. En este tipo de estudios es común rotar los factores o componentes resultantes. Así, Biber (1988) recomienda la rotación promax para permitir cierto grado de correlación entre ellos, y Biber y Egbert (2016) aplican este mismo método rotacional a su propia solución de ACP. En el presente estudio, sin embargo, se ha optado por el ACP no rotado porque, a diferencia de una solución rotada, el ACP no rotado no exige fijar de antemano el número de componentes que se van a extraer, lo que lo hace potencialmente más objetivo (Evert, 2022). Como señala con detalle Le Foll (2024), la decisión sobre cuántos factores o componentes extraer en una solución rotada condiciona también la composición de los componentes resultantes. Así, los dos primeros componentes de una solución rotada de dos CP pueden diferir de los dos primeros CP de una solución rotada de tres (Shakir, 2024). Para evitar este posible sesgo, se retienen los seis primeros componentes de la solución no rotada y se interpretan como tales. Esta decisión tiene además la ventaja de que los componentes resultantes no están correlacionados entre sí, lo que facilita su interpretación, dado que un elevado solapamiento de rasgos entre componentes dificultaría asignarles etiquetas distintivas.

El análisis multidimensional se llevó a cabo con el lenguaje de programación R (R Core Team, 2025), que empleó el paquete psych (Revelle, 2024). Para reducir la dimensionalidad de los conjuntos de datos se utilizó el ACP. La tabla de datos final incluye 74 variables con un KMO de 0,75 (medio) y un 34 % de la varianza global explicada. Las puntuaciones de las dimensiones se han calculado utilizando el método de regresión, predeterminado en el paquete psych[4].

Figura 1

Gráfico de dispersión de los valores propios obtenidos mediante el ACP



La siguiente tabla muestra la estadística descriptiva de los seis CP.

Tabla 2

Datos estadísticos de los componentes principales

CP	Tipo III ANOVA para las tres categorías de traducción	Contraste	Coeficiente de determinación (r^2)
CP1	Categoría no significativa	-----	0.00
CP2	Categoría: $p < 0.001$	TH - TAN: $p < 0.001$, GPT-4 - TAN: $p < 0.001$	0,07
CP3	Categoría: $p < 0.001$	TH - GPT-4: $p < 0.001$, TH - TAN: $p < 0.001$, GPT-4 - TAN: $p < 0.001$	0,64
CP4	Categoría: $p < 0.001$	TH - GPT-4: $p < 0.001$, TH - TAN: $p < 0.001$, GPT-4 - TAN: $p < 0.001$	0,37
CP5	Categoría: $p < 0.001$	TH - GPT-4: $p < 0.001$, TH - TAN: $p < 0.001$	0,06
CP6	Categoría: $p < 0.001$	TH - GPT-4: $p < 0.001$, TH - TAN: $p < 0.001$	0,15

El valor r^2 muestra el coeficiente de determinación, es decir, la proporción de variabilidad que se puede explicar a través de la agrupación de textos o subcategorías (TH, TAN y GPT-4). Como puede verse en la tabla, el CP1 no es significativo ya que las distintas subcategorías establecidas no muestran diferencias. Los CP 3 y 4 tienen un valor r^2 más alto, además de diferencias significativas entre todas las subcategorías. En el caso del CP3, el valor es de 0,64, lo que significa que el CP3 es altamente significativo

y proporciona el 64% de la explicación de las diferencias entre TH, TAN y GPT-4. Aunque el valor del CP4 es inferior al anterior (r^2 0,37, 37%), sigue siendo importante y explicativo, lo que sugiere que esta dimensión también contribuye a distinguir las distintas subcategorías, pero con un impacto algo menor. En CP2, la TAN difiere tanto de TH como de GPT-4. En los CP 5 y 6, la TH difiere de GPT-4 y la TAN. No es posible describir todos los datos en detalle, por lo que nos centraremos en CP3 y CP4, ya que son las dimensiones más potentes para distinguir nuestras subcategorías. La Tabla 3 ofrece los CP resultantes y las características cargadas para estas dimensiones.

Tabla 3

Rasgos lingüísticos de las dimensiones obtenidas del ACP

CP	+/-	Características lingüísticas y pesos
CP3	+	THRCother - *that* relative clauses not preceded by a stance noun (0.68) COMM - Communication verbs (0.56) MDCA - Modal *can* (0.5) VBNRel - Non-finite past participial relative clauses (0.42) NNTECH - Nouns technical (0.38) JJATother - Attributive adjectives having no additional tag (0.38) ThVLIK - *that* subordinate clauses (other than relatives) preceded by likelihood verbs (0.34) MDCO - Modal *could* (0.34) AWL - Average word length (0.33) VBGCls - Non-finite present participial clauses (0.32)
	-	VBGRel - Non-finite present participial relative clauses (-0.31) NNPLACE - Nouns place (-0.31) BEMA - *be* as main verb (-0.33) ACT - Activity verbs (-0.44) RP - Verb particles (-0.5) WHSCother - *WH* complement clauses not preceded by a stance verb (-0.62)
CP4	+	VBG - Non-finite verb *-ing* forms (0.51) ACT - Activity verbs (0.49) NNQUANT - Nouns quantity (0.4) TIME - Time references (0.38) DEMO - Demonstrative pronouns and articles (0.36) VBN - Non-finite *-ed* verbforms (0.36) ASPECT - Aspectual verbs (0.36) OCCUR - Occurrence verbs (0.34) VBNRel - Non-finite past participial relative clauses (0.33) NCOMP - Noun compounds (0.3)
	-	ELAB - Elaborating conjunctions (-0.3) NNCOG - Nouns cognitive (-0.39) PrepNSTNC - Prepositions preceded by stance nouns (-0.43) MDMM - Modals *may* and *might* (-0.46)

4. Resultados

En esta sección se identifican en primer lugar los componentes y se interpretan como dimensiones en función de las funciones comunicativas que comparten. A continuación, se analiza la distribución de los rasgos lingüísticos identificados a través de los tres métodos de traducción (TH, TAN y GPT-4), con ejemplos de texto que

ilustran cada dimensión. Como valor de referencia se estableció, siguiendo a Biber (1988), la cifra de 0,30: solo los rasgos que superan este umbral se interpretan como parte identificativa de cada CP. La carga positiva o negativa no indica relevancia, sino grupos de rasgos que se distribuyen en los textos de forma complementaria. Dado el carácter inicial y exploratorio del estudio, nos centramos en los dos CP con mayor coeficiente de determinación (r^2) —CP3 y CP4—, por ser las dimensiones con mayor poder diferenciador.

4.1 Dimensión 3

Este primer componente cuenta con un total de 16 rasgos lingüísticos, donde 10 ostentan una carga positiva y 6 una carga negativa. Entre los rasgos con cargas positivas más destacadas se encuentran oraciones de relativo (THRCother, 0,68; VBNRel, 0,42), las subordinadas introducidas por verbos de probabilidad (ThVLIK, 0,34) y oraciones precedidas por participio de presente (VBGCLs, 0,32). A estos rasgos se suman el uso de verbos de comunicación (COMM, 0,46), los modales *can* (MDCA, 0,49) y *could* (MDCO 0,33), nominalizaciones que denotan términos especializados (0,38), adjetivos con función atributiva (JJATother, 0,38) y una mayor longitud media de palabra (0,32). En conjunto, estas características reflejan un discurso formal y denso, alineado con las convenciones del registro jurídico, que destaca la precisión terminológica y la complejidad gramatical.

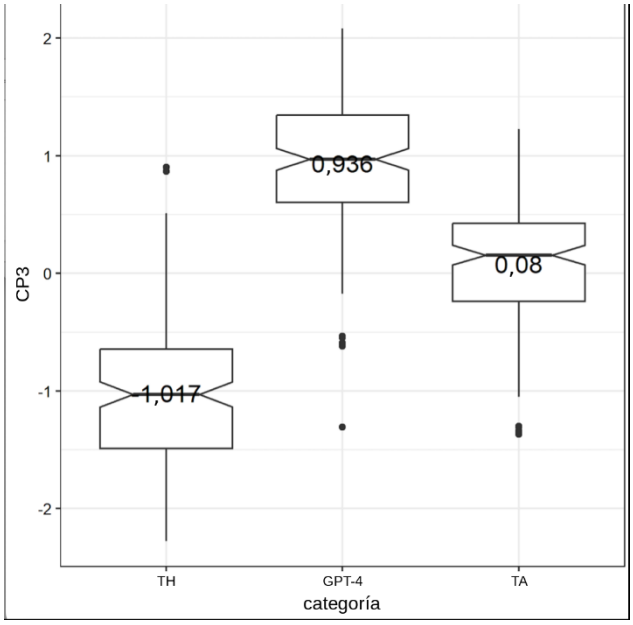
Por el contrario, los rasgos con pesos negativos —como el uso del verbo *be* como núcleo verbal (BEMA, –0,32), los verbos de actividad (ACT, –0,44), las preposiciones (RP,–0,49) y oraciones de relativo precedidas de pronombres wh- (WHSCother–0,62)— señalan a un estilo más directo y orientado a la claridad expositiva.

En cuanto a los valores obtenidos por cada tipo de traducción en esta dimensión, se observa que GPT-4 se sitúa en el extremo positivo (0,936), seguido por la TAN (0,08), que ocupa una posición intermedia, mientras que la TH alcanza un valor negativo notable (–1,017). Esta distribución sugiere que GPT-4 tiende a generar traducciones con un grado elevado de formalidad y tecnicismo, caracterizadas por una mayor densidad sintáctica. La TAN, en cambio, mantiene un perfil estilístico más neutro, posiblemente debido a su entrenamiento centrado en la generalización de estructuras. Por último, la TH muestra una inclinación hacia un estilo más accesible y claro, lo cual puede explicarse por decisiones traductológicas orientadas a facilitar la comprensión del texto jurídico por parte del lector final, sin por ello abandonar los parámetros propios del discurso jurídico.

Esta dimensión pone de relieve el papel del traductor humano como mediador discursivo, capaz de adaptar el estilo sin comprometer la función comunicativa ni la precisión jurídica; podría denominarse *exposición elaborada vs. expresión directa.*

Figura 2

Media el CP3



La Figura 2 muestra la polarización de los métodos de traducción, con la TAN en posición intermedia y neutra. El ejemplo 1 ilustra los rasgos negativos de la dimensión, característicos de la TH y coherentes con el género de las sentencias del TJUE, entre ellos las oraciones de relativo precedidas por wh- o that (VBGRel, WHSCother) y los verbos de acción (ACT) o el verbo to be (BEMA):

Ejemplo 1: Archivo # HT_ECJ_SP_0023_EN

The national court [NNPLACE] decided to refer the following [ACT] questions to the Court of Justice for a preliminary ruling: Should the exemption from taxes payable to the State, province, municipality and other entities of public law which [WHSCother] Spanish law grants to public credit institutions provided [ACT] that they possess the status of taxpayers (Article 29 of Law 13/71 on the organization [NNPLACE] of official credit) be regarded as incompatible with the principle of fair competition [...]

Algunos de estos los resultados obtenidos corroboran los rasgos lingüísticos mencionados por Trklja y McAuliffe (2019), llamados *formulaic metadiscursive devices* en las sentencias emitidas por el TJUE, como es el caso del uso del verbo *to be* como verbo principal (BEMA) en estructuras del tipo *it must be*. El uso de los verbos de acción (ACT) en este caso también apunta a una finalidad metadiscursiva para mostrar transiciones en el discurso (*it follows*). En cuanto al uso de nombres (NNplace), en el caso de este ejemplo, su uso responde, principalmente, al empleo del sustantivo *Court* o el genérico

organization como un objeto de autoridad impersonal a través del uso de la tercera persona (*the national Court decided*).

4.2 Dimensión 4

El CP4 presenta también poder diferenciador entre los tres métodos de traducción, algo menor que en la Dimensión 3 ($r^2 = 0,37$; Tabla 2), pero suficiente para agrupar los textos del corpus.

Esta dimensión se caracteriza por 14 rasgos lingüísticos, donde 10 ostentan un peso positivo y 4 un peso negativo. Los rasgos con cargas positivas en esta dimensión se asocian a una variedad sintáctica p (VBG, 0,51), VBN (0,35) y VBNRel (0,32); una variedad en cuanto a la carga semántica verbal, donde predominan verbos que denotan acción (ACT, 0,48), aspectuales (ASPECT, 0,35) y verbos que describen hechos (OCCUR, 0,33); y una variedad nominal, donde se agrupan nombres que expresan cantidad (NNQUANT, 0,39), demostrativos (DEMO, 0,36) y nombres compuestos (NCOMP, 0,29). Por último, existen también referencias temporales (TEMP, 0,38). Por el contrario, los rasgos con cargas negativas, como los sustantivos cognitivos (NNCOG, -0,39), las conjunciones elaborativas (-0,30), los modales *may* y *might* (MDMM, -0,46), y ciertas construcciones con sustantivos que denotan actitud (-0,42).

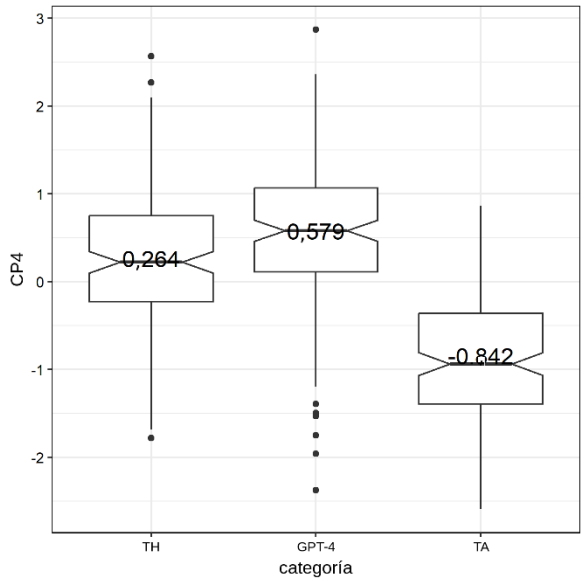
Con ello, los rasgos positivos se relacionan con un estilo procesual, típicamente empleado para describir eventos, secuencias de acciones o etapas en el desarrollo de un procedimiento (Biber, 1988). En el contexto jurídico, estos elementos son fundamentales para construir un discurso que muestre cómo actúa el tribunal, en qué momento lo hace y en qué condiciones procesales, especialmente en los procedimientos del TJUE, donde la función principal del texto es interpretar o valorar la validez del derecho de la UE (Albors-Llorens, 2020). Por el contrario, en el caso de los rasgos lingüísticos con carga negativa, se observa que tienden a densificar el texto, generando un estilo más abstracto, reflexivo y alejado de la acción procesual.

De acuerdo a los resultados obtenidos, las traducciones generadas por GPT-4 presentan un valor notablemente positivo (0,579), lo que indica una fuerte presencia de estructuras procesuales y dinámicas. Esto sugiere que el modelo, probablemente debido a su entrenamiento con grandes volúmenes de textos formales, reproduce un estilo más alineado con la finalidad comunicativa del discurso jurídico: describir procedimientos y acciones jurídicas con precisión secuencial. Las TH, con un valor positivo más moderado (0,264), muestran un equilibrio entre claridad, corrección formal y adecuación al registro, conservando variedad sintáctica y riqueza léxica. En cambio, la TAN muestra un valor claramente negativo (-0,842), lo que indica una mayor densidad conceptual, menos diversidad estructural y una menor orientación hacia la acción. Este patrón refleja la tendencia de los sistemas neuronales de producir textos más homogéneos, literales y conceptuales, menos sensibles al propósito retórico del texto jurídico. En conjunto, estos resultados confirman que tanto la TH como la realizada

por GPT-4 se aproximan más al estilo procesual característico del discurso jurídico institucional que la TAN. En este sentido, podríamos nombrar esta dimensión como orientación procedimental vs. expresión conceptual.

Figura 3

Media para el CP4



La Figura 3 ilustra estos resultados y los ejemplos 2 y 3 dan muestra de lo expuesto. El ejemplo 1 pertenece al subcorpus de GPT-4, donde pueden observarse ejemplos de variedad sintáctica. Por ejemplo, las formas VBG, formas verbales terminadas en -ing, actúan como marcadores discursivos (*following a judgment*) o construcciones de discurso argumentativo (*as meaning that*). Esto sugiere que este tipo de textos suele estar más estructurado y refleja el estilo del lenguaje jurídico de la UE, y más concretamente de las sentencias emitidas por el TJUE. Por otro lado, la variedad verbal centrada en verbos que denotan acciones ACT (*Transport operations carried out under*), el modo en el que se desarrolla una acción (ASPECT, *the decision was being considered*) y ocurren unos hechos (OCCUR, *becomes competent to recover*) refleja una progresión temporal de los eventos descritos en las sentencias. A ellos se unen la variedad nominal (DEMO, *that offence*) como ejemplos de densidad y precisión léxica.

Ejemplo 2: Archivo #IA_ ECJ_SP_0062_EN

Articles 454 and 455 of Regulation No 2454/93 laying down provisions for the implementation of Regulation No 2913/ 92 establishing the Community Customs [NCOMP] Code are to be interpreted as meaning [VBG] that [DEMO], when the presumption that competence to recover a customs debt [NCOMP] lies [ACT] with the Member State in whose territory an offence committed in the course of a TIR

transport [NCOMP] operation was detected is rebutted following [VBG, ACT] a judgment establishing that that [DEMO] offence was committed in the territory of another Member State, the customs authorities [NCOMP] of the latter Member State become [OCCUR] competent to recover that [DEMO] debt, provided [ACT] that the facts giving [ACT] rise to the offence became [OCCUR] the subject of legal proceedings within two years of the date [NNQUANT] on which the guaranteeing [VBG] association for the territory in which the offence was detected was notified thereof.

Por otro lado, el ejemplo 3 refleja los rasgos lingüísticos con carga negativa y asociados, según nuestros resultados, a la TAN. Ello demuestra un mayor nivel de abstracción y estilo argumentativo, con abundancia de sustantivos que denotan cognición (*need, responsability*). En cuanto a las estructuras preposicionales (*in view of, responsability of, the realization of*) refuerzan la estructura impersonal, lógica y jerárquica del discurso jurídico.

Ejemplo 3: Archivo # MT_ECJ_SP_0133_EN

For a decision [NNCOG], agreement or practice to affect trade among member States, it must be assumed to be a sufficient degree of probability [NNCOG], according to a set of objective elements of fact [NNCOG]or law, which have a direct or indirect influence, real or potential, on the flows of exchanges between member States, so as to fear that they might [MDMM] hinder the realization of [PrepNSTNC] a single market among member States.

5. Discusión

En vista de los resultados obtenidos, podemos confirmar nuestra hipótesis de partida: los tres métodos de traducción —TH, TAN y GPT-4— presentan patrones léxico-gramaticales estadísticamente diferenciables (Tabla 2), agrupados en dos componentes con un poder explicativo notable (CP3, $r^2 = 0{,}64$; CP4, $r^2 = 0{,}37$). Este resultado respalda empíricamente la premisa central del análisis multidimensional sobre la coocurrencia sistemática de rasgos léxico-gramaticales agrupables en dimensiones funcionales (Tabla 3), y constituye evidencia directa de que el método de traducción deja una huella lingüística sistemática en el texto. Esta huella es coherente, en un plano temático general, con la noción de efecto de imprimación que Rico Pérez (2025) documenta en textos institucionales del español traducidos mediante TA. Si bien su estudio se centra en una dirección lingüística distinta —hacia el español— y no emplea análisis multidimensional, nuestro trabajo identifica, mediante ACP, los rasgos léxico-gramaticales concretos asociados a esa diferenciación en el inglés y las dimensiones funcionales en las que se organizan.

Nuestro estudio se sitúa, además, en la intersección de dos líneas de investigación que aplican el análisis multidimensional a la traducción institucional, aunque con lógicas

metodológicas distintas. Como Hu et al. (2016), que proponen un modelo estadístico multi-rasgo, de extracción inductiva, para contrastar el inglés escrito traducido y no traducido, nuestro estudio no parte de dimensiones preexistentes, sino que las extrae del propio corpus. A diferencia de Hu et al. (2016), sin embargo, desplazamos el eje de comparación de traducido/no traducido a tres métodos de generación de la traducción (TH, TAN y GPT-4) sobre un mismo conjunto de textos fuente. Por otro lado, nuestro trabajo se relaciona con el de Biel et al. (2025), quienes aplican un nuevo análisis multidimensional a cuatro registros institucionales en polaco, entre ellos las sentencias del TJUE. Este estudio no compara métodos de traducción —se centra en la variación entre registros dentro de una misma lengua—, y confirma que el género que aquí analizamos, las sentencias del TJUE, es susceptible de una caracterización multidimensional rigurosa dentro del discurso jurídico-institucional europeo, lo que refuerza la pertinencia de aplicar esta misma metodología a la variable que nos interesa, el método de traducción.

La relación más directa de nuestro estudio es, sin duda, con Sánchez Ramos y Shakir (2026). El presente estudio no se limita a ampliar aquel diseño mediante la incorporación de GPT-4 como tercera modalidad: la introducción de un tercer término de comparación modifica sustancialmente lo que puede inferirse de la posición que ocupa la TAN en cada dimensión. En una comparación de solo dos modalidades, cualquier polo identificado queda necesariamente definido por la oposición entre ambas, de manera que no es posible determinar si el perfil de la TAN refleja un extremo estilístico genuino o si, simplemente, ocupa esa posición por ausencia de un tercer punto de referencia. Nuestros resultados —obtenidos mediante una extracción de componentes completamente nueva sobre el corpus ampliado, y no mediante la reutilización de la solución factorial de aquel estudio— permiten resolver esta indeterminación a nivel interpretativo, y lo hacen, además, de forma distinta en cada dimensión. En la Dimensión 3 (exposición elaborada vs. expresión directa), GPT-4 ocupa el extremo positivo y la TH el extremo negativo, mientras que la TAN se sitúa en una posición claramente intermedia. Esto sugiere que el perfil de la TAN que pudiera inferirse de comparaciones bidireccionales, como la de Sánchez Ramos y Shakir (2026), no corresponde necesariamente a un polo estilístico propio, sino a una posición moderada dentro del continuum configurado por las tres modalidades analizadas. En la Dimensión 4 (orientación procedimental vs. expresión conceptual), en cambio, el patrón se invierte: la TAN ocupa aquí el extremo negativo, mientras que la TH y GPT-4 convergen, de forma moderada, en el polo positivo. Este contraste —la TAN como término intermedio en una dimensión, pero como término marcado y aislado en la otra— no podía observarse en un diseño de dos modalidades y constituye la aportación empírica específica de ampliar la comparación a tres métodos de traducción: no se trata solo de añadir un tercer perfil lingüístico, sino de mostrar que la posición relativa de la TAN respecto a la TH depende de la dimensión considerada y de qué otras modalidades entran en la comparación, un matiz con implicaciones directas para cualquier estudio

que infiera las propiedades del machine translationese a partir de comparaciones exclusivamente binarias.

La posición intermedia de la TAN en la Dimensión 3 es, además, congruente con la hipótesis del lenguaje restringido de Kruger y Van Rooy (2016): su perfil menos marcado coincide con los rasgos de simplificación y homogeneización que estos autores atribuyen a los textos producidos bajo contacto lingüístico. Sin embargo, la marcada polarización entre TH y GPT-4 —los dos extremos de la dimensión— matiza la hipótesis en un punto relevante: en el registro jurídico-institucional del TJUE, ni la TH ni GPT-4 se ajustan al patrón de lenguaje restringido en esta dimensión. El extremo de expresión directa que ocupa la TH admite una lectura compatible con una mediación discursiva orientada a la claridad y a la adecuación comunicativa para el lector del texto jurídico, más que con una producción limitada. El extremo opuesto de GPT-4, por su parte, es interpretable como un ajuste marcado a las convenciones formales del género, antes que como restricción.

El perfil de la TAN en la Dimensión 4 —el extremo negativo, asociado a un estilo conceptualmente denso, pero escasamente procedimental, con menor variedad sintáctica y nominal— es interpretable a la luz de tres estudios previos sobre las limitaciones léxico-estructurales de la TAN, cada uno de los cuales aporta una pieza distinta de esta interpretación. Frankenberg-Garcia (2022) documenta una inconsistencia léxica considerable en la TAN; nuestros resultados, obtenidos en un par de lenguas y un género completamente distinto —español-inglés, discurso jurídico-institucional—, muestran un patrón convergente (menor variedad sintáctica y nominal en la TAN), lo que sugiere que esta limitación no es un artefacto de un par de lenguas o un dominio concreto, sino un rasgo más general del comportamiento de estos sistemas. Niu y Jiang (2024) señalan la simplificación como principal efecto de imprimación de la TA; nuestro hallazgo es compatible con esta conclusión, aunque no idéntico: la simplificación que documentan se refiere a la reducción de la diversidad léxica, mientras que nuestra Dimensión 4 capta, de forma más específica, una menor orientación hacia la acción y el procedimiento. Ambos fenómenos pueden entenderse como manifestaciones relacionadas de una misma tendencia general de la TAN hacia la homogeneización estructural, pero no deben presentarse como el mismo hallazgo. Por último, la pérdida y el deterioro de la riqueza lingüística que cuantifican Vanmassenhove et al. (2019) ofrece una descripción que se ajusta de forma directa al perfil observado: la TAN de nuestro corpus no solo se aleja del estilo procedimental de TH y GPT-4, sino que lo hace precisamente mediante los mecanismos —menor variedad sintáctica, menor variedad nominal— que estos autores asocian a esa pérdida de riqueza. Estas limitaciones no son solo léxico-estructurales, sino también funcionales: la TAN revela dificultades para adaptarse a los fines retóricos del discurso jurídico coherentes con las que Killman (2023) documenta, mediante una metodología distinta —evaluación de calidad, no análisis multidimensional—, en la traducción automática, neuronal y

estadística, de los rasgos morfosintácticos de sentencias jurídicas en español, el mismo tipo textual analizado aquí.

El perfil de GPT-4 —el extremo positivo tanto en la Dimensión 3 como, de forma más moderada, en la Dimensión 4— tiene un paralelismo directo en Berber-Sardinha (2024), quien, en un diseño confirmatorio, documenta diferencias sistemáticas entre textos generados por IA y textos de autoría humana, con un grado de discriminabilidad notable entre ambos. Nuestros resultados son coherentes con esa tendencia general: GPT-4 ocupa en ambas dimensiones una posición claramente diferenciada, y no intermedia, respecto a TH y TAN. La comparación, no obstante, se limita a este nivel general, ya que ni las dimensiones concretas ni la dirección de la divergencia son equiparables entre ambos estudios, que difieren en dominio, lengua y diseño factorial.

Distinta es la relación con Brivá-Iglesias et al. (2024). Se trata de un tipo de evidencia distinto al nuestro, pues aquel estudio mide calidad de la traducción mediante métricas automáticas y el presente trabajo describe el perfil lingüístico-funcional del texto producido. Que GPT-4 obtenga resultados competitivos en métricas automáticas de calidad en el ámbito jurídico y que, en nuestro estudio, sus traducciones se ajusten con notable precisión a las convenciones léxico-gramaticales de este discurso resultan observaciones compatibles entre sí, si bien proceden de instrumentos y niveles de análisis distintos. Estas lecturas, situadas en el conjunto de la literatura revisada, deben matizarse a la luz de las limitaciones metodológicas del presente diseño, que se detallan a continuación.

En primer lugar, existe una limitación de diseño que incide sobre la interpretación de los resultados correspondientes a la subcategoría de TH. El corpus de traducciones humanas abarca el período 1990-2014, mientras que las traducciones de MTUOC y GPT-4 se generaron en 2025 a partir de los mismos textos fuente en español. Esta asimetría temporal implica que los patrones léxico-gramaticales observados en la TH reflejan las prácticas traductoras propias de ese período, que no tienen por qué coincidir con las actuales. El presente estudio no tenía como objetivo estimar un efecto causal puro del método de traducción con control de todas las variables históricas, sino comparar el comportamiento lingüístico de tres modalidades aplicadas exactamente al mismo conjunto de textos fuente. En ese marco, la comparación es metodológicamente válida para el corpus disponible; sin embargo, cualquier posible efecto diacrónico sobre los patrones de la TH debe entenderse como una limitación del diseño que acota el alcance de las inferencias. Este tipo de asimetría es, por otra parte, inherente a los estudios retrospectivos de corpus que incorporan traducciones humanas auténticas producidas en distintos contextos históricos, ya que la obtención de traducciones humanas contemporáneas para los mismos textos fuente requeriría un diseño prospectivo específico. Como argumento complementario, cabe señalar que el género de las sentencias del TJUE presenta convenciones redaccionales especialmente estables

a lo largo del tiempo, lo que puede atenuar —aunque no eliminar— la magnitud del efecto diacrónico (Biel, 2014; Trklja & McAuliffe, 2019).

En segundo lugar, una limitación adicional afecta a la interpretabilidad de las diferencias observadas entre MTUOC y GPT-4. Estos dos sistemas no difieren únicamente en su arquitectura, sino también, y de forma fundamental, en los datos con los que fueron entrenados: MTUOC fue entrenado específicamente con el Corpus Paralelo de las Naciones Unidas (Ziemski et al., 2016), un corpus de dominio concreto y volumen acotado, mientras que GPT-4 fue entrenado con datos masivos de acceso propietario cuya composición no se conoce públicamente. Es importante delimitar el objeto real de este estudio: no se comparan arquitecturas de traducción en condiciones experimentales controladas, sino el comportamiento lingüístico observable de tres sistemas de traducción reales. Los datos de entrenamiento forman parte de la identidad de cada sistema y, por tanto, del objeto mismo de la comparación; las diferencias observadas describen cómo se comportan estos sistemas tal como existen y se utilizan en la práctica. Ahora bien, esta característica del diseño implica que los resultados no permiten atribuir de forma aislada las diferencias observadas a la arquitectura del sistema, ya que arquitectura y datos de entrenamiento constituyen variables inseparables en el presente diseño. Esta es una limitación importante que debe tenerse en cuenta al interpretar los hallazgos. Futuros trabajos que comparen sistemas entrenados sobre datos equivalentes en dominio y volumen permitirían aislar con mayor rigor el efecto específico de la arquitectura.

Por último, el ACP aplicado al corpus explica, asimismo, un 34% de la varianza total, valor que debe interpretarse en el contexto de estudios multidimensionales aplicados a datos lingüísticos naturales, donde la variación suele distribuirse entre múltiples dimensiones y no se espera que un número reducido de componentes capture la totalidad de la variación lingüística (Biber, 1988; Egbert & Staples, 2019). La varianza residual refleja la complejidad inherente de la variación lingüística en textos naturales y puede estar influida por factores no contemplados en el diseño del presente estudio; identificar esas fuentes de variación constituye una línea de investigación futura que permitiría delimitar con mayor precisión el alcance explicativo del modelo.

CONCLUSIONES

La presente investigación ha ofrecido un estudio que aplica el análisis multidimensional (Biber, 1988) a tres modalidades de traducción en el ámbito jurídico: TH, TAN y GPT-4. A través de un ACP, se han identificado seis dimensiones lingüísticas, de las cuales cinco mostraron un grado de significancia relevante. Este trabajo se ha centrado especialmente en las Dimensiones 3 y 4 por ser las que presentan un mayor poder diferenciador entre los métodos de traducción.

Los resultados confirman que los tres métodos de traducción generan textos con rasgos lingüísticos claramente diferenciables. La TH mantiene un equilibrio entre

precisión formal y claridad comunicativa, adaptándose con eficacia al estilo del género jurídico. La TAN, por su parte, tiende a una producción más neutral, con menor variedad sintáctica y un estilo más abstracto, probablemente como resultado de su diseño centrado en la generalización estructural. En cambio, las traducciones generadas por GPT-4 exhiben una notable complejidad sintáctica y una orientación procesual alineada con los fines comunicativos del discurso jurídico, lo que apunta a un potencial significativo de los modelos de IA generativa en contextos especializados.

En nuestra opinión, el trabajo presenta implicaciones desde distinto ángulos, como el investigador, el pedagógico y el profesional. De un lado, la aproximación multidimensional al estudio de la variación lingüística en distintos métodos de traducción abre nuevas posibilidades de metodología investigadora en el ámbito de la TA. Por otro lado, existen implicaciones pedagógicas para la didáctica de la traducción en cuanto a las tendencias estilísticas y estructurales de los diferentes métodos de traducción en contextos especializados. Por ejemplo, los resultados del análisis multidimensional pueden ayudar a guiar al estudiantado a un uso responsable y efectivo de cada tecnología, especialmente en contextos especializados como el jurídico. Por último, creemos que el análisis multidimensional también tiene implicaciones importantes para los proveedores de servicios lingüísticos que trabajan en el entrenamiento de motores neuronales o de grandes modelos de lenguaje masivo, como el caso de GPT-4, favoreciendo, de este modo, desarrollos más precisos y adecuados al género discursivo de destino. No obstante, dado el carácter exploratorio del trabajo, cabe recordar sus principales limitaciones, ya discutidas en el apartado anterior: el ámbito acotado a un género (sentencias del TJUE), la exclusión de textos poseditados, la sensibilidad de los resultados al diseño concreto del prompt empleado, la asimetría temporal entre el corpus de TH (1990-2014) y las traducciones automáticas y generativas (2025), la imposibilidad de aislar el efecto de la arquitectura del sistema de los datos de entrenamiento en la comparación MTUOC-GPT-4, y la varianza explicada por el ACP (34 %). Todas ellas apuntan a líneas de investigación futura —ampliación a otros géneros, inclusión de textos poseditados, réplica con distintos diseños de prompt y modelos, y comparaciones con traducciones humanas contemporáneas o sistemas entrenados sobre datos equivalentes— que permitirían consolidar y extender los hallazgos aquí presentados.

En definitiva, el presente trabajo sienta las bases para incorporar el análisis multidimensional como una herramienta de análisis rigurosa en los estudios de traducción. Su aplicación no solo permite diferenciar con precisión los rasgos lingüísticos de distintas modalidades traductoras, sino que ofrece un marco interpretativo robusto para comprender las transformaciones del discurso mediadas por la tecnología. Lejos de ser una técnica exclusivamente cuantitativa, el análisis multidimensional se revela como una vía prometedora para explorar el estilo, la función y la calidad de la traducción en la era de la IA.

REFERENCIAS BIBLIOGRÁFICAS

Albors-Llorens, A. (2020). Judicial protection before the Court of Justice of the European Union. En C. Barnardm, & S. Peers (Eds.), *European Union Law* (pp. 283-333). Oxford University Press.

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. En G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233-252). John Benjamins.

Berber-Sardinha, T. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, *4*(1), Article 100083.

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Biber, D. (1995a). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge University Press.

Biber, D. (1995b). On the role of computational, statistical, and interpretive techniques in multi-dimensional analyses of register variation: A reply to Watson. *Text: Interdisciplinary Journal for the Study of Discourse, 15*(3), 341-370.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins.

Biber, D., & Egbert, J. (2016). Register variation on the searchable web: A multi-dimensional analysis. *Journal of English Linguistics, 44*(2), 95-137. https://doi.org/10.1177/0075424216628955

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

Biel, Ł. (2014). *Lost in the eurofog: The textual fit of translated law*. Peter Lang. https://doi.org/10.3726/978-3-653-03986-3

Biel, Ł., Wasilewska, K., & Koźbiał, D. (2025). Dimensions of variation across institutional legal and administrative registers: An MDA analysis of the Polish Eurolect and the national variety. *International Journal of Corpus Linguistics*. https://doi.org/10.1075/ijcl.25126.bie

Brivá-Iglesias, V., Dogru, G., & Cavalheiro Camargo, J. L. (2024). Large language models ad referendum: How good are they at machine translation in the legal domain? *MonTi Monografías De Traducción E Interpretación, 16*, 75-107.

Calzada Pérez, M., & Sánchez Ramos, M. M. (2022). MDA analysis of translated and non-translated parliamentary discourse. En M. Ji, & M. P. Oakes (Eds.), *Corpus exploration of lexis and discourse in translation* (pp. 26-55). Routledge. https://doi.org/10.4324/9781003102694-2

Castilho, S., & Resende, N. (2022). Post-editese in literary translations. *Information*, *13*(2), Article 66. https://doi.org/10.3390/info13020066

Chou, I., & Liu, K. (2024). Style in speech and narration of two English translations of *Hongloumeng*: A corpus-based multidimensional study. *Target*, *36*(1), 77-111.

De Sutter, G., & Lefer, M. A. (2020). On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives*, *28*(1), 1-23.

Egbert, J., & Staples, S. (2019). Doing multidimensional analysis in SPSS, SAS and R. En T. Berber-Sardinha, & M. V. Pinto (Eds.), *Multidimensional analysis research methods and current issues* (pp. 125-144). Bloomsbury.

Evert, S. (2022). *A multivariate approach to linguistic variation*. https://stephanie-evert.de/SIGIL/sigil_R/materials/07_multivar_talk.handout.pdf

Frankenberg-Garcia, A. (2022). Can a corpus-driven lexical analysis of human and machine translation unveil discourse features that set them apart? *Target*, *34*(2), 278-308.

Giampieri, P. (2025). AI-Powered contracts: A critical analysis. *International Journal of Semiotics Law, 38*, 403-420.

Góngora-Goloubintseff, J. G. (2026). Technology-driven changes in the UK legal translation market: A practice-based study of NMT and its impact on freelance translators' workflow. *Perspectives. Studies in Translation Theory and Practice.* https://doi.org/10.1080/0907676X.2026.2684285

Heiss, C., & Soffritti, M. (2018). DeepL traduttore e didattica della traduzione dall'italiano in tedesco. alcune valutazioni preliminari. En L. Anderson, L. Gavioli, & F. Zanettin (Eds.), *InTRAlinea. Special Issue: 'Translation and Interpreting for Language Learners' (TAIL).* https://www.intralinea.org/specials/article/2294

Hu, X., Xiao, R., & Hardie, A. (2016). How do English translations differ from non-translated English writings? A multi-feature statistical model for linguistic variation analysis. *Corpus Linguistics and Linguistic Theory, 15*(2), 347-382.

Ilisei, I., & Inkpen, D. (2011). Translationese traits in Romanian newspapers: A machine learning approach. *International Journal of Computational Linguistics and Applications*, *2*(2), 319-332.

Jiao, W., Wang, W., Huang, J.-T., Wang, X., Shi, S., & Tu, Z. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv*. https://doi.org/10.48550/arXiv.2301.08745

Killman, J. (2023). Rendering morphosyntactic features of legal Spanish judgments using NMT and SMT. En J. Zhao, D. Li, & V. L. C. Lei (Eds.), *New advances in legal translation and interpreting* (pp. 221-242). Springer.

Kruger, H., & Van Rooy, B. (2016). Constrained language. A multidimensional analysis of translated English and a non-native indigenised variety of English. *English World-Wide, 37*(1), 26-57.

Krüger, R. (2020). Explicitation in neural machine translation. *Across Languages and Cultures, 21*(2), 195-216.

Lapshinova-Koltunski, E. (2015). Variation in translation: evidence from corpora. En C. Fantinuoli, & F. Zanettin (Eds.), *New directions in corpus-based translation studies* (pp. 79-99). Language Science Press.

Le Foll, E. (2022). *Textbook English: A corpus-based analysis of the language of EFL textbooks used in secondary schools in France, Germany and Spain* [Tesis de doctorado]. University of Osnabrück. https://doi.org/10.48693/278

Le Foll, E. (2024). *Textbook English: A multi-dimensional approach*. John Benjamins. https://doi.org/10.1075/scl.116

Le Foll, E., & Shakir, M. (2023). *MFTE Python* [Computer software]. https://github.com/mshakirDr/MFTE

Le Foll, E., & Shakir, M. (2024). The Multi-Feature Tagger of English (MFTE): Rationale, description and evaluation. *Research in Corpus Linguistics, 13(2)*, 63-93.

Neumann, S., & Evert, S. (2021). A register variation perspective on varieties of English. En E. Seoane & D. Biber (Eds.), *Corpus-based approaches to register variation* (pp. 143-178). John Benjamins.

Niu, J., & Jiang, Y. (2024). Does simplification hold true for machine translations? A corpus-based analysis of lexical diversity in text varieties across genres. *Humanities and Social Sciences Communications, 11*, 1-10.

Oliver, A. (2021). *MTUOC-Translator* [Computer software]. https://github.com/aoliverg/MTUOC-translator

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2024). *psych: Procedures for psychological, psychometric, and personality research* (Version 2.4.6.26) [Computer software]. Northwestern University. https://cran.r-project.org/package=psych

Rico Pérez, C. (2025). Estudio del efecto de imprimación de la traducción automática sobre un corpus de textos del español institucional. *Revista de Humanidades Digitales, 10*, 48-72.

Rico Pérez, C., Sánchez Ramos, M. M., & Oliver, A. (2020). INMIGRA3: Building a case for NGOs and NMT. En A. Martins, H. Moniz, S. Fumega, B. Martins, F. Batista, L. Coheur, C. Parra, I. Trancoso, M. Turchi, A. Bisazza, J. Moorkens, A. Guerberof, M. Nurminen, L. Marg, & M. Forcada (Eds.), *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation* (pp. 469-471). European Association for Machine Translation.

Sánchez Ramos, M. D. M., & Shakir, M. (2026). A corpus-based approach to human and machine translation: a multidimensional analysis of judgments of the Court of Justice of the European Union. *Perspectives*, 1-21. https://doi.org/10.1080/0907676X.2026.2687796

Shakir, M. (2024). An exploratory investigation of functional variation in South Asian online Englishes. *English Language & Linguistics, 28*(2), 371-400. https://doi.org/10.1017/S1360674324000017

Toral, A. (2019). Post-editese: An exacerbated translationese. En M. Forcada, A. Way, B. Haddow, & R. Sennrich (Eds.), *Proceedings of Machine Translation Summit XVII: Research track* (pp. 273-281). European Association for Machine Translation.

Trklja, A., & McAuliffe, K. (2019). Formulaic metadiscursive signalling devices in judgments of the Court of Justice of the European Union: A new corpus-based model for studying discourse relations of texts. *International Journal of Speech, Language and the Law*, *26*(1), 21-55.

Vanmassenhove, E., Shterionov, D., & Way, A. (2019). *Lost in translation: Loss and decay of linguistic richness in machine translation* [Ponencia]. Machine Translation Summit XVII, Dublín, Irlanda.

Vigier Moreno, F., & Sánchez Ramos, M. M. (2017). Using parallel corpora to study the translation of legal-system bound terms: The case of names of English and Spanish Courts. En R. Mitkov (Ed.), *Computational and corpus-based phraseology: Second International Conference, Europhras 2017, London, UK, November 13-14, 2017 proceedings* (pp. 260-273). Springer.

Volansky, V., Ordan, N., & Wintner, S. (2015). On the features of translationese. *Digital Scholarship in the Humanities*, *30*(1), 98-118.

Yates, S. (2006). Scaling the tower of Babel Fish: An analysis of the machine translation of legal information. *Law Library Journal, 98*(3), 481-500.

Ziemski, M., Junczys-Dowmunt, M., & Pouliquen, B. (2016). The United Nations parallel corpus v1.0. En N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the tenth international conference on language resources and evaluation (LREC'16)* (pp. 3530-3534). European Language Resources Association. https://doi.org/10.63317/263tmvnpzevv

NOTAS

[1] El presente trabajo se ha desarrollado en el marco del proyecto "La variación de registro en traducción automática: un análisis multidimensional aplicado a un corpus en el tercer sector social" (PID2024-157849OB-I00), financiado por MICIU /AEI /10.13039/501100011033 / FEDER, UE.

[2] La asimetría temporal con respecto a las traducciones de MTUOC y GPT-4 generadas en 2025 a partir de los mismos textos fuente, así como la generada por la distinta procedencia de los datos de entrenamiento de MTUOC y GPT-4 se abordan en la sección Discusión.

[3] Los autores desean expresar su agradecimiento a Antoni Oliver (Universitat Oberta Catalunya) por su inestimable ayuda con el sistema de traducción automática neuronal MTUOC y a Sergi Álvarez Vidal (Universitat Autònoma de Barcelona) por su colaboración en el acceso y uso de la API de GPT-4.

[4] Los datos obtenidos están disponibles en el siguiente repositorio:

https://osf.io/4cgx7/overview?view_only=3f8910c28730442ab5939a4f36c8e8a9